# From Scratch to Smart

A 3-Phase Demo: Pre-training, SFT, and RLHF

# The "Baby" Model: Our Starting Point

" We start with a `ToyLLM`. Its 'brain' (weights) is completely random. When we ask it a question, its answer is meaningless garbage. "

Prompt: 'alice followed'
Response: 'alice rabbit queen cat curious...'

# The Core NLP Building Blocks

## Tokenization

A "vocabulary" to turn words like 'alice' into simple numbers like [12].

## Embeddings

A "dictionary" that maps these numbers to *meaning vectors* that the model can learn.

# Phase 1: Foundational Pre-training

## Goal: Learn Language Structure

We teach the model the basics of language by making it predict the next word over and over.

- **Method:** Next-Word Prediction

- **Data:** Simple sentences from "Alice in Wonderland."

- **Example:** Show it "alice followed the..." and teach it to predict "...white".

# Pre-training Result: A "Completer"

" The model is no longer random! It completes
sentences based on the 'Alice' facts it learned.
But it can't *answer* questions. "

Prompt: 'who did alice...'
Response: 'who did alice follow the...'

# Phase 2: Supervised Fine-Tuning (SFT)

## Goal: Teach Factual Answers

We teach the model to be a helpful "expert" by showing it "flashcards" of (Question, Answer) pairs.

- **Method:** Instruction Following

- **Data:** Q&A pairs.

- **Example:**
  Prompt: `'who did alice follow'`
  Answer: `'alice followed the white rabbit'`

# SFT Result: A "By-the-Book" Expert

## Good: Trained Questions

It perfectly answers the *specific questions* it was trained on.

Prompt: 'what did the cat do'

Response: 'the cheshire cat smiled'

## Bad: New Questions

When asked a *new* question, it falls back on its pre-trained "completer" knowledge.

Prompt: 'who was alice'

Response: 'alice was very curious'

# Phase 3: Reinforcement Learning (RLHF)

## Goal: Teach Preferences

We teach the model what a "good" answer *feels* like, even for new questions.

1. Model generates an answer.

2. A "Reward Model" (Judge) scores it.

3. We update the model to get more "points" from the judge in the future.

# RLHF Result: An "Aligned" Model

"

The model is now "steered" by our preferences. Our judge "rewarded" the answer `'alice was very curious'` , teaching the model that this is a good, helpful response.

"

Prompt: 'who was alice'
Response: `'alice was very curious'`

# The Model's Evolution: A Comparison

## Pre-Trained

**Prompt:** 'who did alice follow'
**Response:** 'who did alice follow the...'

**Prompt:** 'who was alice'
**Response:** 'who was alice said the...'

## SFT

**Prompt:** 'who did alice follow'
**Response:** 'alice followed the white rabbit'

**Prompt:** 'who was alice'
**Response:** 'alice was very curious'

## RLHF

**Prompt:** 'who did alice follow'
**Response:** 'alice followed the white rabbit'

**Prompt:** 'who was alice'
**Response:** 'alice was very curious'

# Key Demo Takeaways

💡 **Pre-training** teaches *language* and *facts*. It creates a "completer."

🎓 **SFT** teaches *instruction following*. It creates an "expert."

👍 **RLHF** teaches *preferences*. It creates an "aligned" assistant.

# Questions?

# Image Sources



https://media.istockphoto.com/id/1412637153/photo/magic-book-with-open-pages-and-abstract-lights-shining-in-darkness-literature-and-fairytale.jpg?s=612x612&w=0&k=20&c=rc3Bh74ePX8QPGnOi5oeEYOwFaJYsseXu8A0dGFqRzI=

Source: www.istockphoto.com



https://png.pngtree.com/png-clipart/20250422/original/pngtree-stylized-q-a-speech-bubbles-illustration-png-image_20849901.png

Source: pngtree.com



https://ecdn.teacherspayteachers.com/thumbitem/Thumbs-Check-for-Understanding-1609884359/original-268414-1.jpg

Source: www.teacherspayteachers.com